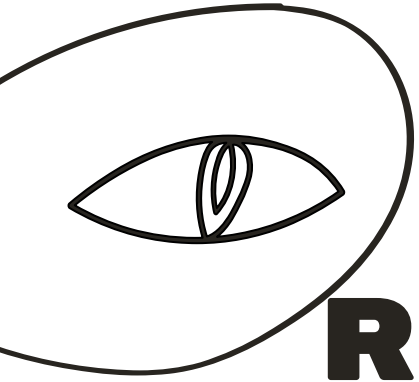
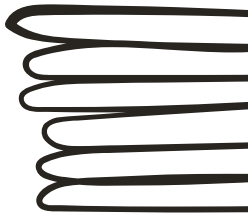
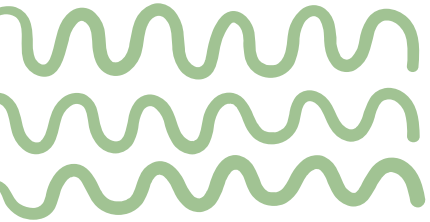




RESEARCH PAPER ABSTRACT NLP

Prepared By:

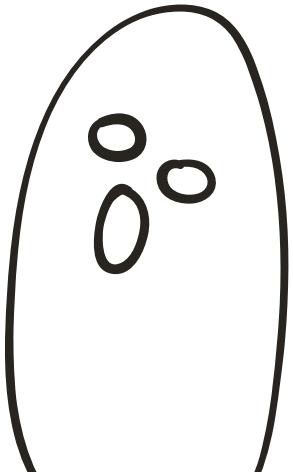
Taef AlKhales
Sara Hawi



INTRODUCTION

An abstract is a short summary of your completed research. It is intended to describe your work without going into detail.

A lot of time during research can be spent looking for relevant or similar papers to a particular paper you found. Most of the time, reading the abstract alone can tell you whether this paper is useful and relevant.



80K+ ABSTRACTS

Abstracts from research papers about different fields :
Statistics, Physics, Math & Computer Science

kaggle

arXiv

ABSTRACT

This paper proposes a methodology to detect early signs of Parkinson's disease (PD) through free-speech in uncontrolled background conditions. The early detection mechanism uses signal and speech processing techniques integrated with machine learning algorithms. Three distinct speech databases containing patients' recordings at different stages of the PD are used for estimation of the parameters during the training and evaluation stages. The results reveal the potential in using Random Forest (RF) or Support Vector Machine (SVM) techniques. Once tuned, these algorithms provide a reliable computational method for estimating the presence of PD with a very high accuracy.

PRE-PROCESSING

1- CONVERT ABSTRACT TO LOWER CASE

2- REMOVING DIGITS AND PUNCTUATION

3- TOKENIZING BY WORD

4- REMOVING STOP WORDS

5- LEMMATIZATION

VECTORIZATION

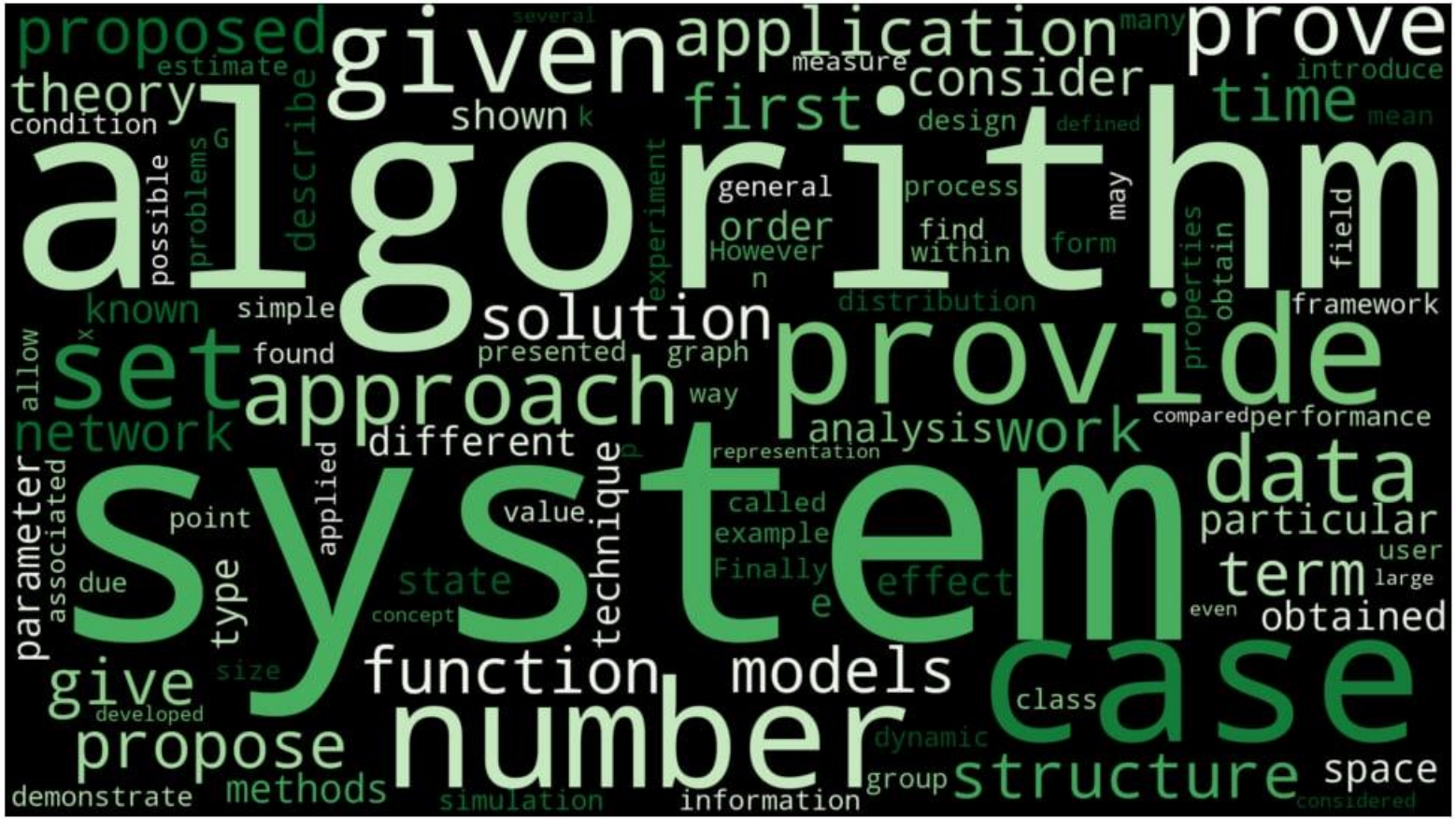
COUNT VECTORIZER

Gives more weight to high frequent words like : **paper, propose, system**

TF-IDF

Better results, gives unique words a higher weight like: **wave**

100 most common words in Papers Abstracts



TOPIC MODELING

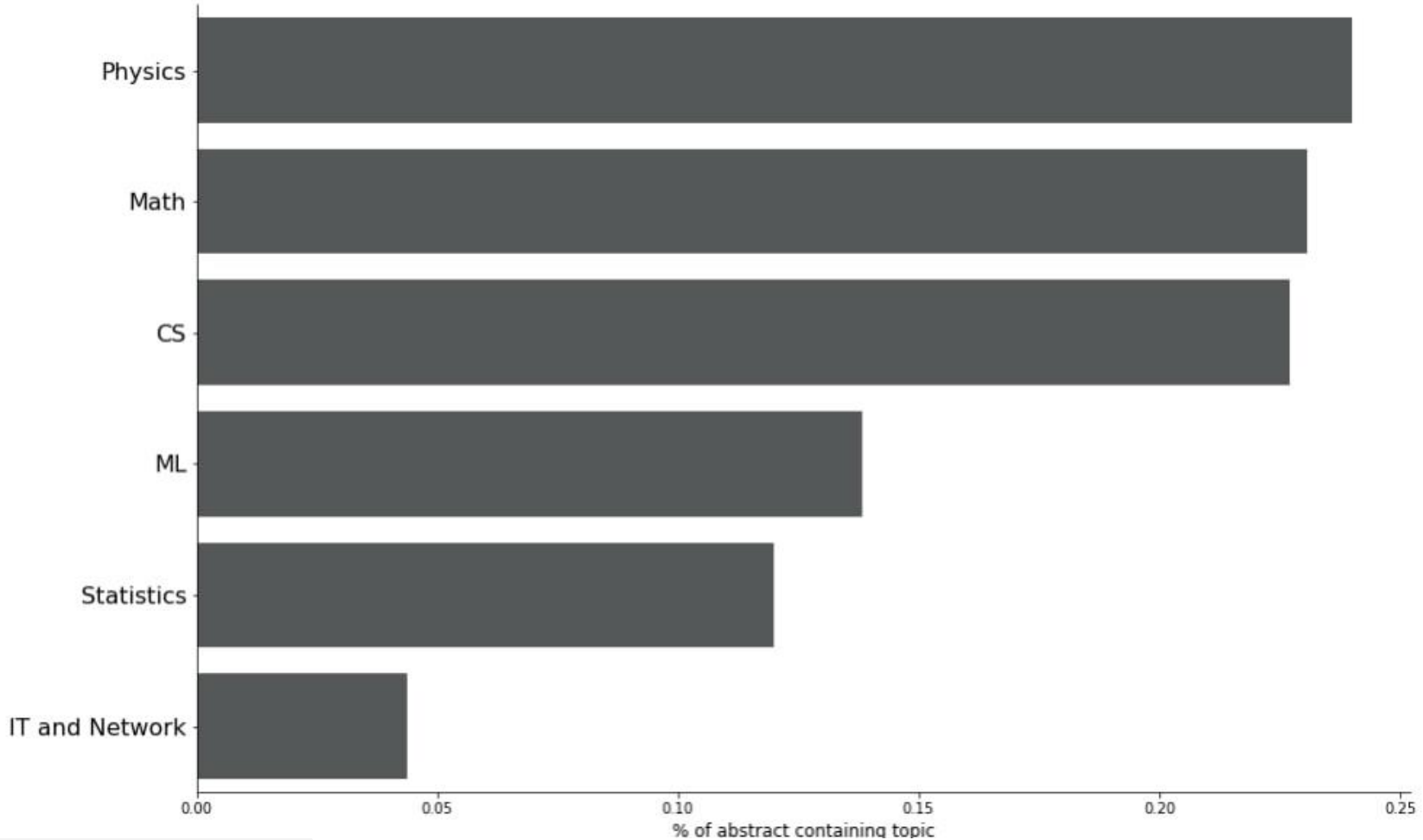
LATENT DIRICHLET ALLOCATION
COMPONENTS = 4

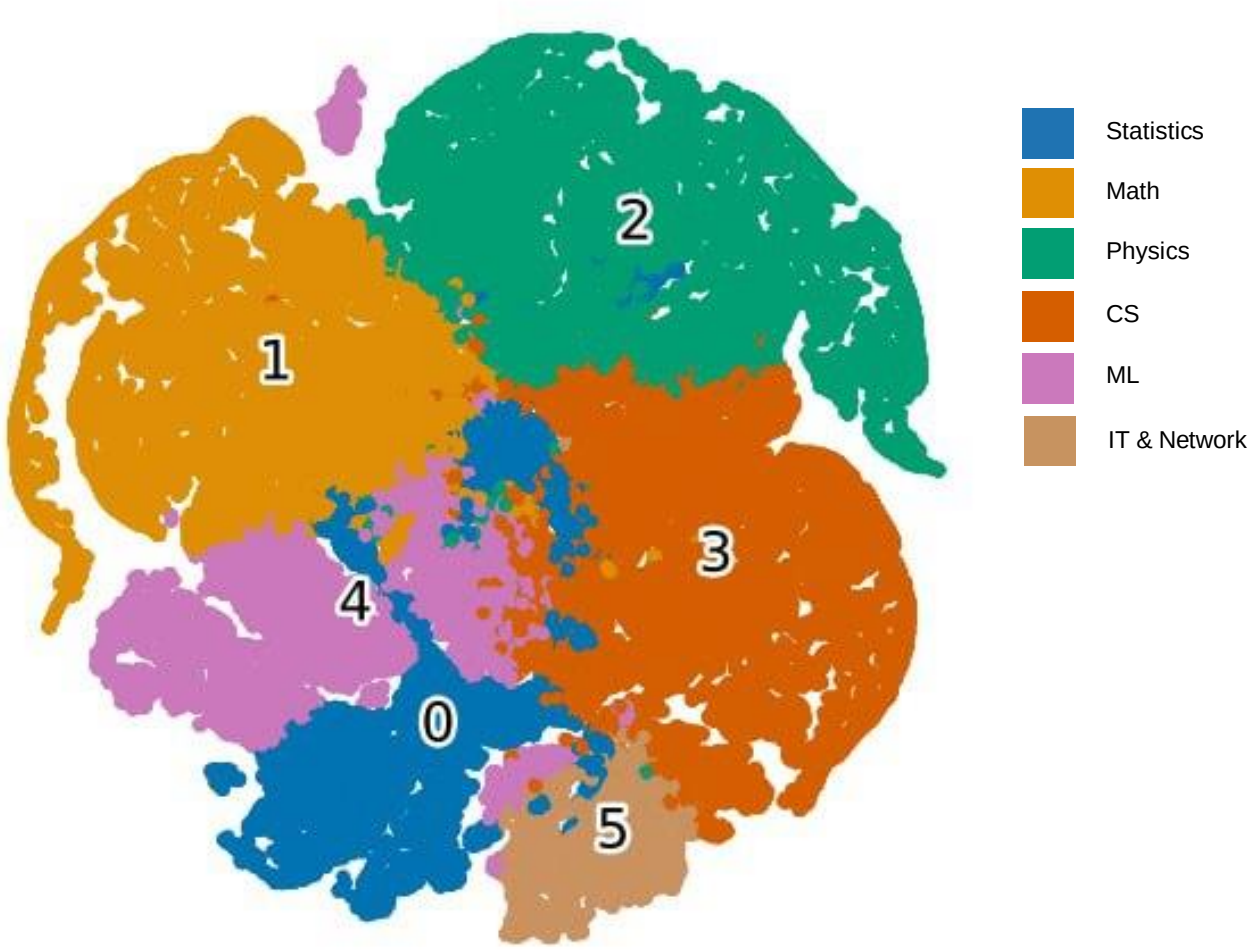
NON-NEGATIVE MATRIX FACTORIZATION
COMPONENTS = 6

LDA RESULTS DEMO

NMF RESULTS

STATISTICS	MATH	PHYSICS	COMPUTER SCIENCE	MACHINE LEARNING	INFORMATION TECHNOLOGY
distribution	algebra	field	system	algorithm	network
function	space	energy	data	bound	node
data	prove	wave	application	solution	social
estimation	finite	equation	language	complexity	sensor
sample	theorem	electron	software	polynomial	wireless
likelihood	function	particle	design	approximation	protocol
bayesian	let	beam	web	optimization	routing
probability	complex	magnetic	program	tree	traffic
gaussian	property	laser	tool	learning	packet
density	representation	quantum	development	linear	security


Physics
Math
CS
ML
Statistics
IT and Network
0.00
0.05
0.10
0.15
0.20
0.25
% of abstract containing topic


2
1
3
4
0
5
Statistics
Math
Physics
CS
ML
IT & Network

NMF RESULTS ON ABSTRACTS - IT & NETWORKS

We consider a set of k autonomous robots that are endowed with visibility sensors (but that are otherwise unable to communicate) and motion actuators. Those robots must collaborate to reach a single vertex that is unknown beforehand, and to remain there hereafter. Previous works on gathering in ring-shaped networks suggest that there exists a tradeoff between the size of the set of potential initial configurations, and the power of the sensing capabilities of the robots (i.e. the larger the initial configuration set, the most powerful the sensor needs to be). We prove that there is no such trade off. We propose a gathering protocol for an odd number of robots in a ring-shaped network that allows symmetric but not periodic configurations as initial configurations, Yet, uses only local weak multiplicity detection. Robots are assumed to be anonymous and oblivious, and the execution model is the non-atomic CORDA model with asynchronous fair scheduling. Our protocol allows the largest set of initial configurations (with respect to impossibility results) yet uses the weakest multiplicity detector to date. The time complexity of our protocol is O(n2), where n denotes the size of the ring. Compared to previous work that also uses local weak multiplicity detection, we do not have the constraint that k < n/2 (here, we simply have 2 < k < n - 3).

STATISTICS	MATH	PHYSICS	COMPUTER SCIENCE	MACHINE LEARNING	INFORMATION TECHNOLOGY
0.3%	0.8%	0.4%	0.66%	1.967%	2.548%

NMF RESULTS ON ABSTRACTS - PHYSICS

The Fermilab Linac delivers a variable intensity, 400-MeV beam to the The MuCool Test Area experimental hall via a beam line specifically designed to facilitate measurements of the Linac beam emittance and properties. A 10 m, dispersion-free and magnet-free straight utilizes an upstream quadrupole focusing triplet in combination with the necessary in-straight beam diagnostics to fully characterize the transverse beam properties. Since the Linac does not produce a strictly elliptical phase space, tomography must be performed on the profile data to retrieve the actual particle distribution in phase space. This is achieved by rotating the phase space distribution using different waist focusing conditions of the upstream triplet and performing a de-convolution of the profile data. Preliminary measurements using this diagnostic section are reported here.

STATISTICS	MATH	PHYSICS	COMPUTER SCIENCE	MACHINE LEARNING	INFORMATION TECHNOLOGY
0.84%	0.285%	3.29%	0.4%	0%	0%

RECOMMENDATION SYSTEM DEMO

(CONTENT BASED)

CONCLUSIONS

- Using NMF, more categories were successfully and accurately identified (ML, IT & Networks)
- The performance of the model can be improved by using supervised techniques.
- Add more diversity to the dataset (medical, psychology, environmental, political, literature, etc.)
- The performance of the recommendation system can be enhanced by adding the index of the respective paper to show the most popular and useful papers in a certain field.

THANK YOU FOR LISTENING!